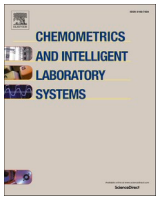
Chemometrics and Intelligent Laboratory Systems 250 (2024) 105152

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Virtual biopsies for breast cancer using MCR-ALS perfusion-based biomarkers and double cross-validation PLS-DA

E. Aguado-Sarrió [a,b,*], J.M. Prats-Montalbán [a,b], J. Camps-Herrero [c], A. Ferrer [a,b]

[a] Kenko Imalytics S.L, Calle Pedro Duque, Edificio 9B, Universitat Politècnica de València (UPV), Valencia, Spain
[b] Grupo de Ingeniería Estadística Multivariante (GIEM), Universitat Politècnica de València (UPV), Valencia, Spain
[c] Grupo Ribera Salud, Avda. Cortes Valencianas, 58, 46015, Valencia, Spain

A B S T R A C T

Functional MRI is, currently, the most sensitive technique in breast cancer for detecting early tumors, and perfusion (DCE-MRI) has become the most important sequence to depict and characterize angiogenesis and neovascularization. In this work, we propose the use of new biomarkers that are related to clear physiological phenomena, obtained from MCR-ALS as an alternative to curve-based pseudo-biomarkers and pharmacokinetics models. In order to provide a discrimination and prediction model between healthy tissue and cancer, we propose using PLS-DA with double cross-validation (2CV) and variable selection, repeated several times and obtaining excellent average results for the performance indexes (f-score: 0.9149, MCC: 0.8538, AUROC: 0.8794). After selecting the optimal prediction model, a unique probabilistic map called "virtual biopsy" that shows in different colors the probability that each pixel of the image has a tumor behavior is obtained, helping the specialist with the identification and characterization of breast tumors with only one easy-to-interpret biomarker map.

1. Introduction

Breast cancer is one of the most prevalent cancers at present, being the main cause of cancer mortality within the female population. There were approximately 2.3 million new breast cancer cases worldwide in 2020 [1,2]. Due to the lack of objective indicators of cancer presence (risk of developing tumoral tissue) and the high number of deaths that is consequence of this fact, breast cancer has been the focus of attention of the health community for a long time, addressing a large part of the efforts to the early diagnosis of the disease and prevention. The survival rate of breast cancer is really high if they are detected in very early stages, dropping dramatically to less than 15 % when detected in advanced stages [3].

Due to the high incidence of this disease, the efforts of the health community should be focused in the development of better diagnostic and treatment techniques. For this reason, different imaging techniques have been developed, including functional magnetic resonance imaging (MRI) due to its ability to early detect angiogenesis and neovascularization, main indicators of a tumor process [4].

Angiogenesis and neovascularization are biological processes associated to tissues with increased oxygen and nutrients demands. These processes seldom occur in healthy subjects, but they are strongly present in pathological conditions such as tumors. The formation of these new and tortuous vessels produces an increase in the blood perfusion, which can be studied with dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) [5]. In DCE-MRI studies, an exogenous contrast media is administered intravenously and diffuses from the capillary network into the extravascular extracellular space (EES) and returns, establishing a dynamic relationship between the image signal intensity changes and the amount of contrast media that passes and diffuses into a certain tissue. The capability to analyze tumor angiogenesis in a quantitative and reproducible way from DCE-MR images has important applications to depict and grade tumors, and also to early evaluate the therapeutic response after treatment onset [6,7]. To obtain quantitative measurements, it is necessary to fit and characterize the intensity versus time curves associated to each voxel of the image. Out of the different approaches proposed to achieve this characterization, mathematical pharmacokinetic models have become the most popular way due to their ability to provide clinically-oriented biomarkers in tumor analysis. Nevertheless, when the temporal resolution is low (which is the case of breast DCE-MRI, where only 6 different time points are acquired), the pharmacokinetic analysis is not useful and the radiologists prefer easier indicators such as the characteristics of the curves vs. time (slope, time to maximum, signal drop from peak percentage, etc.). However, this methodology can be improved with the application of multivariate image analysis (MIA) techniques to this kind of sequences. Recently, new biomarkers obtained from multivariate curve resolution (MCR) models have been also proposed [8,9] in prostate cancer studies in order

* Corresponding author. Grupo de Ingeniería Estadística Multivariante (GIEM), Universitat Politècnica de València (UPV), Valencia, Spain.
E-mail address: eric@kenkoimalytics.es (E. Aguado-Sarrió).

https://doi.org/10.1016/j.chemolab.2024.105152
Received 18 September 2023; Received in revised form 14 April 2024; Accepted 27 May 2024
Available online 28 May 2024

to provide more easy-to-interpret biomarkers, associated to specific physiological phenomena like angiogenesis.

The aims of this paper are the following:

1. To provide new parametric maps obtained from the MCR imaging biomarkers helping the radiologists with new tools for breast cancer diagnosis. Additionally, in this work, three new biomarkers have been introduced in order to help the professionals in the visualization of the lesions in the parametric maps.
2. To test these biomarkers for classification/prediction of healthy vs. lesion tissues. The MCR biomarkers are introduced in a classification model using partial least squares discriminant analysis (PLS-DA) technique [10,11], studying the sensitivity and specificity associated to the rate of false positives (FP) and false negatives (FN), and testing different well-known figures-of-merit like f-score, MCC, AUROC, specificity, sensitivity and FN/FP percentages.
3. To provide a new feature as a unique probabilistic map called "Virtual biopsy", which is a probability heat map obtained after obtaining the optimal PLS-DA model. These virtual biopsies are projections of new external cases onto the classification model, obtaining a cancer probability value for each pixel of the new image.

The paper is organized as follows. In Section 2 the data base is described and the multivariate statistical methods used in this work are introduced. Section 3 presents the results of the MCR model, showing their capability for breast cancer classification/prediction. Examples of virtual biopsies are also provided for testing the model in real patients, discussing the pros and cons of using this kind of biomarkers. Finally, Section 4 provides the conclusions.

2. Material and methods

2.1. Patient database

The database for the PLS-DA method consists of 25 histologically-confirmed cases of breast tumors. DCE-MRI sequences were acquired with 3T Philips Medical Systems equipment in all cases, ensuring full breast coverage (720 images total, 400 kb per DICOM image, 288 Mb total per case approximately) per case, 120 slices acquired at 6 different dynamics, in-plane resolution of 432 × 432 voxels, each one measuring 0.8734 × 0.8734 × 2 mm3). Slice thickness 1.9950 mm. Spacing between slices: 1.5 mm. repetition time 5.7454 s, echo time 3,15 s. Flip angle: 18 Acquisition time: 13 min approx. No preprocessing is required to the sequences since they are obtained under the same configuration and MRI equipment.

Then, 60 additional cases with the same sequence characteristics have been used as an external set for virtual biopsy testing. These cases are also histologically-confirmed cases of breast tumors.

Both data bases are composed of ductal non special type (NST) infiltrating carcinomas; the tumoral subtypes considered in this study are described in Table 1:

Reference tumor regions of interest (ROIs) were identified and confirmed by pathological anatomy and the "ground truth" about the lesions is well known. Then, the ROIs were segmented by an expert radiologist in breast imaging and her team (up to five radiologists), considering image findings and then, biopsy location, using a structured reporting scheme for evaluating breast cancer known as breast imaging – reporting and data system (BI-RADS®) [12]. This scale ranges from 1 to 6, based on the description of the lesion findings, and allows defining two different types of ROIs:

- DL: Dominant Lesion (malignant), related to well-known cancer tissue (BI-RADS = 6).
- H: Healthy, related to healthy tissue (normal breast) (BI-RADS = 1).

These ROIs are manually defined by the radiologists, considering the BI-RADS score and the biopsy result, and are used as the gold reference for tissue classification. It is assumed that the behavior of the healthy regions is similar in cancer and healthy patients. Therefore, it could be used safely as the healthy tissue reference. The PLS-DA **X** matrix is built after selecting 7276 pixels (rows) from 25 different breast cancer patients (3498 pixels with lesion and 3778 healthy pixels).

All patients gave consent for using their medical images, which were anonymized before post-processing. The local Ethics Committee approved the study protocol.

Table 1
Description of the different types of tumoral subtype considered in this work.

Tumoral Subtype	Abb.	description
Luminal A	LUM A	This group includes tumors that are estrogen receptor positive and progesterone receptor positive but human epidermal growth factor receptor 2 negative.
Luminal B	LUM B	This type includes tumors that are estrogen receptor positive, progesterone receptor negative, and human epidermal growth factor receptor 2 positive.
Human epidermal growth factor receptor 2	HER 2	This type includes tumors that are negative for estrogen receptor and progesterone receptor, but positive for human epidermal growth factor receptor 2.
Triple Negative	TN	This type, which is also called the "basal type," includes tumors that are negative for the estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2. They are the most aggressive.

2.2. MCR-ALS models calculation

One possible way to look for physiological meaningful dynamics is by applying multivariate statistical projection models to the DCE-MRI data. When dealing with images, the application of these models is known as multivariate image analysis (MIA) [13,14], usually based on principal component analysis (PCA) [15,16]. Nevertheless, a relevant drawback of the application of PCA in DCE-MR image analysis is that the estimated dynamics patterns are orthogonal by design. The orthogonality of the principal components is a limitation to model different perfusion behaviors that are not necessarily orthogonal. Therefore, PCA can never be a good approach when it comes to revealing the actual underlying physicochemical or biological phenomena from DCE-MRI, since it aggregates new orthogonal components according to their explained variance, losing their direct interpretability and, as a consequence, the information provided by the "scores-maps" (instead of the distribution maps) incorporates noise or mixes different phenomena; hence lowering the performance. In order to overcome these drawbacks, it is possible to use more flexible models as multivariate curve resolution-alternating least squares (MCR-ALS) [17,18], which do not impose this constraint. MCR is preferred to PCA because of its ability to provide physiologically interpretable behaviors by imposing a priori knowledge on the model. MCR-ALS is an iterative method that performs a bilinear decomposition of an **S** matrix by means of an alternating least squares optimization algorithm,

$$\mathbf{S} = \mathbf{C}(\mathbf{D})^{\mathbf{T}} + \mathbf{E} \quad \text{(eq. 1)}$$

Where **S** contains the signal intensity registered for each voxel in rows; $\mathbf{D}^{\mathbf{T}}$ is a matrix containing in its rows each of the perfusion dynamic behaviors modelled; **C** gathers in its rows the relative contribution of each dynamic behavior modelled at each voxel of the image; and **E** is a residual matrix [8,9]. A scheme about the dimensions of the matrices in eq. (1) is shown in Fig. 1.

In order to apply the MCR algorithm, the **S** matrix is built by unfolding the n_1 x n_2 voxels of the 120 slices of each case that form a grayscale image for each channel in the time dimension (T). A scheme of

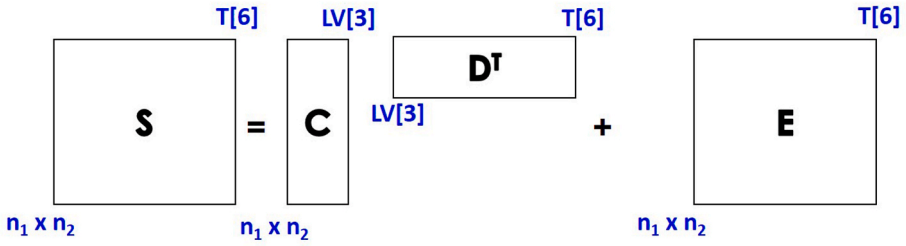


Fig. 1. MCR model schema containing the dimensions of the matrices for the DCE-MRI perfusion analysis. The 3 components (LV) considered in this model are the dynamic behaviors related to NT (normal tissue), VT (vascularized tissue) and CMA (contrast media arrival).

the unfolding process needed to obtain the **S** matrix from the perfusion images is shown in Fig. 2.

One MCR model is obtained per case. During the MCR-ALS iterative process, equal length normalization is applied to the columns of **D** matrix containing the dynamic behaviors modelled (Fig. 3). This way the values of the relative contribution of each dynamic behavior modelled (columns of **C** matrix) can be directly compared with each other. Once the iterative process is completed, by refolding each column of **C** matrix into the original spatial dimensions, new biomarker images are obtained (Fig. 4), which allow locating those voxels more related to each one of the corresponding modelled dynamic behaviours.

Note that MCR-ALS is based on an iterative process that can provide infinite solutions for the same data matrix causing a problem known as ambiguity in the solution [17,18]. This problem can be solved by imposing constraints commonly related to prior knowledge about the problem faced. So, it is possible to obtain easier-to-interpret solutions, which also tend to be unique when constraints introduced under the hypothesized assumptions are accurate.

For this, two additional constraints were imposed:

- Non-negativity on the pixel relative contribution values **C**, because the intensity in a pixel has to be nonnegative.
- Non-negativity on the dynamic profiles $\mathbf{D^T}$.

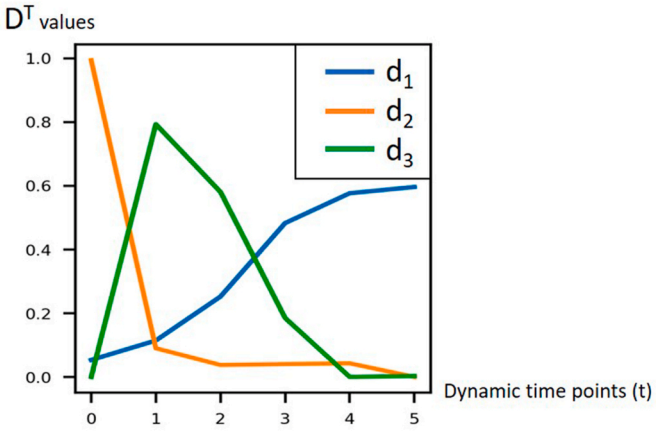


Fig. 3. Dynamic behaviors obtained from MCR-ALS. $\mathbf{d_1}$: type NT, normal tissue; $\mathbf{d_2}$: type CMA, contrast media arrival; $\mathbf{d_3}$: type VT, vascular tissue.

According to these papers [8,9] and specific clinical advice about the breast cancer behaviour, two components are considered in MCR models regarding to the dynamic patterns related to perfusion studies that can

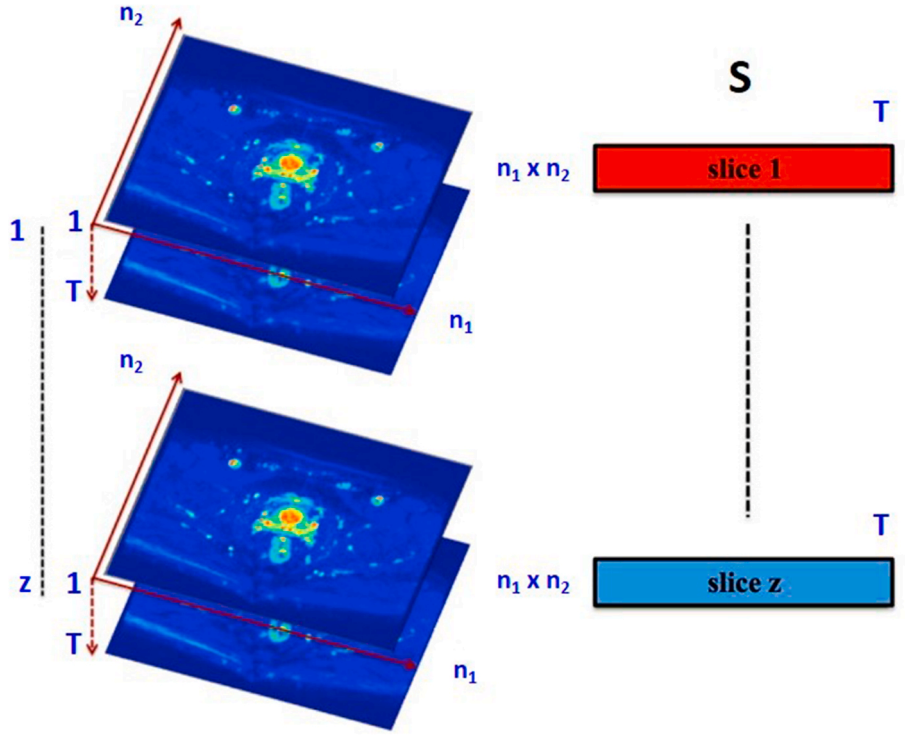


Fig. 2. Data structure for MCR analysis. Images are n_1 x n_2 pixels size, T is the number of time points (perfusion dynamics), z is the number of slices. The 2D unfolded **S** matrix dimensions are ((n_1 x n_2 x z)) x T.

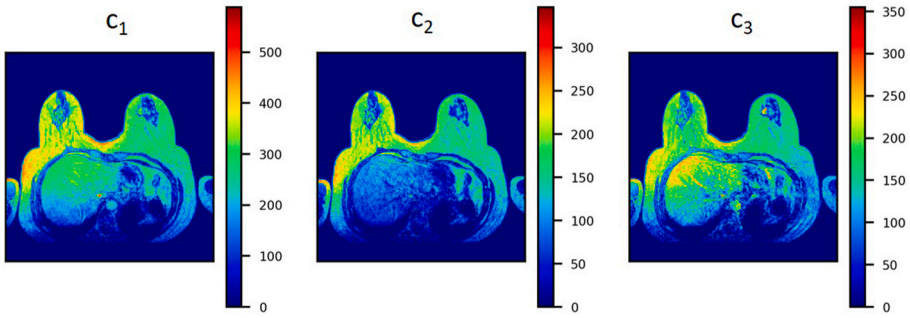


Fig. 4. Parametric distribution maps of MCR-ALS imaging biomarkers (C_1, C_2, C_3).

be expected using the "a priori" knowledge about the process. (i.e., normal tissue (type NT (d_1)) and vascular tissue (type VT (d_3)) (Fig. 3). Also, a third component is obtained because of the appearance of an artifact introduced by the MRI equipment overestimating the signal intensity before the contrast arrival, defined as CMA (d_2) [8,9]).

Furthermore, from the three parametric maps obtained from the MCR-ALS model and clinical advice, three new imaging biomarkers are calculated related to the differences between the previous dynamics (Fig. 5). These biomarkers have been proposed in order to improve the visualization of the lesions in the parametric maps. This is a result of applying the MCR-based model to the breast and a solid way of removing the "base" intensity of the perfusion MRI artifact, consisting on a pre-contrast arrival to the tissue, because the vascularized tissue does not have any contribution in C_2 but it does in C_3 and C_1. This way, the visualization is *clearly* improved as can be seen by comparing Figs. 4 and 5.

$$B1 = C_1 - C_2\ ;\ 2 = C_3 - C_2\ ;\ B3 = \frac{C_1 + C_3}{2} - C_2 \qquad \text{(eq. 2)}$$

These imaging biomarkers are proposed in this paper since they clearly improve the visualization of the tumor regions. Also, they will be included in the classification model for testing their prediction ability. Moreover, also in this case, it is possible to include the residual sum of squares (RSS) of the MCR model (sum of squares of rows of residual matrix **E** in eq. (1)) as an additional potential biomarker measuring the lack of fit in each voxel location.

2.3. Variable selection using double cross-validation (2CV) with PLS-DA

In this paper, the capability of imaging biomarkers obtained from MCR-ALS have been tested for breast cancer classification using latent-variables based statistical classification methods, such as partial least squares-discriminant analysis (PLS-DA) [10,11]. Seven MCR biomarkers are proposed for discrimination:

- C_1 (NT behavior)
- C_2 (CMA behavior)
- C_3 (VT behavior)
- B1
- B2
- B3
- RSS (Residual sum of squares obtained from MCR model)

A 7-column **X** matrix is constructed by stacking all the selected voxels in rows with the value of the seven biomarkers in columns. The selected voxels are the ones defined by the radiologists as DL (lesion) or H (healthy) ROIs for each patient. These ROIs are logical images, each one associated to a specific slice of the breast DCE-MRI sequence. Thus, the **X** matrix is constructed by stacking all the voxels assigned at the corresponding local image for all the patients. For the same voxels, a 2-column **Y** matrix is defined with two dummy variables (0–1). The first column defines the "DL" variable (value 1 if the voxel belongs to the "DL" region, and value 0 if otherwise). The second column defines the "H" variable and is built complementary to the first one. Once fitted the PLS-DA model, the class showing a predicted higher value is assigned to each corresponding voxel. Using these PLS-DA class assignments, the assigned category of the voxels is compared to the original classification. Therefore, if the voxel belongs to H, it can be evaluated as a TN (True Negative) or FP (False Positive) depending on its prediction, and as a TP (True Positive) or FN (False Negative) if it is DL. This yields the confusion matrix from which different classification performance indexes are calculated:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{recall} = \text{sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \end{aligned} \qquad \text{(eq. 3)}$$

These indexes are combined in three new performance indexes, f-score, Matthews correlation coefficient (MCC) and the area under the receiver operating characteristic (AUROC), chosen to evaluate the overall classification model performance. The f-score [19] is defined as the weighted harmonic mean of precision and recall:

$$\text{f-score} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \qquad \text{(eq. 4)}$$

It ranges between 0 and 1, and takes the maximum when the

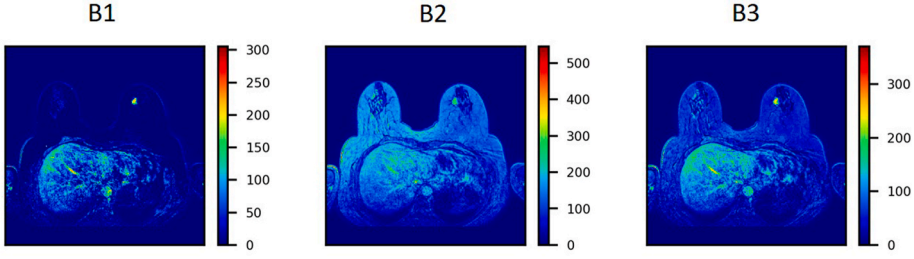


Fig. 5. Parametric maps of the alternative MCR-ALS imaging biomarkers: B1 (left), B2 (middle), B3 (right).

precision and recall are one (the number of FP and FN is zero). The closer the f-score is to one, the better the model is in terms of classification performance. Additionally, the Matthews correlation coefficient [20] (MCC) is the correlation coefficient between the observed and predicted binary classifications; it returns a value between −1 and + 1. A coefficient of +1 represents a perfect prediction, 0 corresponds to no better than random prediction, and −1 indicates total disagreement between prediction and observation:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad \text{(eq. 5)}$$

The AUROC is defined as the area under the curve in the sensitivity vs (1-specificity) plot where a value of 0.5 is equivalent to random guess and 1 is perfect classification.

The variable selection method proposed in this work (Fig. 6) is a wrapped double cross-validation (2CV) with variable selection, showing high similarities with other 2CV methods [19,20]. This method consists of dividing the voxels from the cases of the data set (25 patients in this work) in three randomized groups of cases (i.e., 9, 8, 8 patients), defined as training, validation and test. All the voxels from each case have their own class identification (DL or H) and are located at their corresponding group. This way, the voxels from a specific case are always included in the same group in order to avoid any type of overfitting. Moreover, the number of voxels in each category has been balanced as much as possible in order to avoid any bias of the model.

The method performs as follows [21]: For each iteration (It) the data set is split in the 3 different sets (training, validation and test) and then the sets are permuted 3 times (P) obtaining 3 different data arrangement. For each one of these data arrangements, starting from a number of latent variables (NLV) equal to one, the training set is used for PLS-DA model building, using all the variables (biomarkers) considered. Then, projecting the validation set onto the model fitted with the training set, an initial f-score (0) is calculated evaluating the performance in the model classification.

The f-score (0) is stored and then, the values of the $\mathbf{B_{PLS}}$ coefficients for each variable or biomarker are compared with their "null" distribution obtained after breaking the correlation structure between **X** and **Y** of the training set. This breakage process consists of randomizing the order of the **Y** matrix rows keeping the same **X** and building a PLS-DA model to obtain the "null" model coefficients. This is internally repeated 500 times in order to obtain the null distribution [19,20] of these coefficients. This way, the variable (biomarker) is removed from the **X** matrix if the coefficient of a certain variable is not statistically significant. It is considered statistically significant if the real coefficient is out of the central 95 % range for the random null distribution values (i.e. $\alpha = 0.05$).

At each iteration step, all the non-statistically significant variables are removed all at once and, in the next step, a new PLS-DA model is built with the remaining statistically significant variables from the training set **X** matrix, obtaining a new value of f-score (1) after projecting again the same validation set (only using the remaining statistically significant variables) onto the new PLS-DA model. If the new f-score (1) is higher than f-score (0), the previous model is discarded (maintaining the new model) and the new value of f-score (1) is updated. In this case, the iterative process continues with a new variable selection comparing the new $\mathbf{B_{PLS}}$ coefficients after breaking again the correlation structure between **X** and **Y**. However, if f-score (1) is lower than f-score (0), the best model is the one considered in the previous step. This iterative process is repeated until the f-score (n) is lower than the one obtained in the previous step, f-score (n-1), keeping this "best" model with its associated f-score (n-1) and its own variable selection. From the best model selected, a final external set (test) is projected onto this model

```
For It =1:500
    Do X_train, X_val, X_test splitting
    For P = 1:3
        Do the permutation between sets (not needed in P=1)
        For NLV = 1:(N-1)
            initialize f-score (prev.)
            Do PLS-DA model with X_train
            Predict X_val and obtain new f-score (Validation set)
            While f-score(new) > f-score (prev.)
                estimate B_PLS from null distribution
                Remove the non-significant variables and rebuild X
                f-score (prev.) = f-score (new)
                do PLS-DA model with the new X_train
                Predict X_val and obtain new f-score (Validation set)
                f-score (new) > f-score(prev.)?
            End
            Store last model as best model for NLV
            Predict with X_test and obtain f-score (test)
        End
        Compare the best models of each NLV and select the final model
    End
    Arrange the 3 final model results in the results matrix
End
```

Fig. 6. PLS-DA with 2CV and variable selection method pseudocode. The algorithm pipeline is represented with loop structure and indentation for better understanding. It (500) represents the number of times the data set is split. P (3) is the number of set permutations between training, validation and test for each It. N is the maximum number of variables in **X** (7). NLV is the current number of latent variables.

obtaining the "final f-score". This value is stored for further comparisons.

Afterwards, the NLV is increased by one repeating the same process explained for one LV. This way, at the end we loop through N-1 (where N is the initial number of variables considered in the **X** matrix) NLV obtaining N-1 models (with its own variable selection) and final f-score. Although the new biomarkers (B1, B2 and B3) are linear combinations of the original MCR biomarkers, the iterative model will always reach the maximum NLV (N-1) permitted since it is not computationally demanding and in some iterations, the optimal model is obtained with 6 LV (the maximum number allowed). It should be noted that if the number of variables remaining in **X** is higher than the current NLV, the variable selection loop stops, the previous model is stored, and the loop continues with the next NLV. After this, the final f-score of these N-1 models are compared and the highest final f-score determines the best model with the best variable selection. Note that if different models provide the same value of final f-score, the most parsimonious model is preferred for simplicity.

The initial group randomization is repeated 500 times obtaining 1500 (500 × 3) different data arrangements. Variable selection [19–21] is applied in each iteration and for each NLV tested as long as the f-score improves in order to determine which biomarkers supply relevant information for classification, selecting them for obtaining the optimal model. This step is added to the iterative process because when variable selection is not applied to PLS-DA, a suboptimal model can be obtained, lowering the classification performance. In this work, all the variables were almost always selected in the final model (above 95 % inclusion) but these results cannot be known "a priori". This high selection rate is a confirmation that every variable from MCR is important for the discrimination between classes (even the linear combinations of the original biomarkers) and their inclusion improves the goodness of classification. A pseudocode of the algorithm is shown in Fig. 6.

At the end of the process, the value of the final f-score, MCC, the variable selection (biomarkers that remain in the model), the percentage of TN, TP, FN, FP and the related number of latent variables (NLV) for the "best" model obtained for each arrangement of the 1500 groups are obtained. This way, afterwards, it is possible to select the most relevant biomarkers from those appearing with the highest frequency in the 1500 final models, as well as computing the overall final f-score, MCC, AUROC, TN, TP, FN and FP values. At the end of the process, the overall best model is selected from the best models of this arrangement by averaging the $\mathbf{B_{PLS}}$ coefficients of the 10 best models since the differences between them are not statistically significant (results not shown). However, the best model out of the 1500 iterations can be used as well obtaining almost identical results when calculating the virtual biopsies in section 3.2. Both options have been tested obtaining equivalent virtual biopsy images (results of the comparison not shown).

Once the methods are described and in order to ease the understanding of the results, a summarized step-by-step of the proposed methodology is presented:

1 Read the MRI perfusion images and build the data matrices structure following the unfolding process described in Fig. 2Then, stack all the slices of the same case and build a unique 2D matrix with pixels in rows and time points in columns able to be analyzed by MCR.
2 Calculate the MCR-ALS algorithm applying non-negativity constraints for **D** and **C**, and compute 3 components in the initialization (PCA can be used to obtain a good starting point, when no a priori knowledge is available).
3 Once **C** is obtained, calculate the 3 new biomarkers with simple arithmetic operations and rearrange the values of them in the original spatial resolution to obtain the parametric maps of the biomarkers.
4 Select the pixels of the lesions (indicated by the group of expert radiologists) and built the database described in 2.3 to apply PLS-DA with variable selection.
5 Repeat the PLS-DA iterative method enough times (1500 in this case) to obtain an optimal model with the correspondent $\mathbf{B_{PLS}}$ coefficients
6 Project external cases (i.e., not included in the 2CV iterative method) onto the classification model, obtaining a cancer probability value for each pixel of the image. Rearrange the new probability map in the original spatial resolution to obtain the virtual biopsy parametric maps.

3. Results and discussion

First, the results obtained from the PLS-DA model are analyzed, selecting the optimized model in terms of the performance indexes described in the last section (f-score, MCC, AUROC, sensitivity, specificity and FN/FP percentages). Then, this model is used for prediction purposes, obtaining a probabilistic map of the breast in order to differentiate between tumoral and healthy tissue, which is called "virtual biopsy". Some examples are shown in the last point of this section.

3.1. Optimized PLS-DA model selection

The results repeating the optimization process 1500 times are shown in Tables 2 and 3:

As can be seen in the tables, the results are extremely good, obtaining high levels of f-score (0.9149 mean value), MCC (0.8538 mean value) and AUROC (0.8794 mean value), which means the optimized model is good for classification purposes, and better than other studies [21] (f-score = 0.7857). These results can also be analyzed in terms of the false positive rate (1.3 % mean rate) and the false negatives rate (6.14 % mean rate). This low FP rate can be assumed close to 0 since the model is built at pixel level and better than other studies [21] (7.5 %). Although the FN rate in slightly higher, it is still low and better than other studies [21] (13.5 %), although reference 21 is referred to another type of cancer (prostate). It should be noted that, as indicated, this analysis is performed at the pixel level, where the classification is more difficult than at the ROI level.

Regarding the number of NLV, the average obtained is 3.37 and the best model has 2 NLV. However, 300 out of 1500 iterations reached the maximum number of NLV (6).

In order to show a better visual understanding of the results, some histograms of the principal indexes are presented in Fig. 7:

When the f-score distributions for the training and test sets are compared (Fig. 7, top), the differences are low, being the training f-score slightly better, as it is assumed because it is obtained from the set used to build the model. But the distributions are mostly overlapped, showing good prediction consistency. Similar results (not shown) are obtained for MCC and AUROC.

Regarding the variable selection, the next table shows the results of the percentage of inclusion of each biomarker:

As shown in Table 4, the biomarkers proposed in this paper are almost always included in the model (inclusion rate higher than 95 %). This result demonstrates the importance of every biomarker for classification, even the ones obtained from the original MCR imaging biomarkers, and the lack-of-fit represented by the RSS. All of them are needed for improving the classification performance. However, it must be noted that this result might be not always the case: it depends on the

Table 2
Results of the iterative process, after 1500 iterations (f-score, MCC, AUROC, specificity, sensitivity and precision).

	f-score	MCC	AUROC	specificity	sensitivity	precision
Mean	0.9149	0.8538	0.8794	0.9686	0.8958	0.9754
Median (prt 50)	0.9147	0.8533	0,9176	0.9719	0.8938	0.9780
prt 25	0.8917	0.8155	0.8747	0.9583	0.8681	0.9676
prt 10	0.8749	0.7894	0.7560	0.9432	0.8487	0.9550

Table 3
Results of the iterative process, after 1500 iterations (FP(%) and FN(%)).

	FP(%)	FN(%)
Mean	1.3112	6.1440
Median (50th percentile)	1.1865	6.2044
75th percentile	1.7290	7.8896
90th percentile	2.3798	9.1773

quality of the biomarkers applied and their relationship with the organ analyzed. This model is also tumor-type dependent; it works better in highly vascularized lesions such as the ones of the data base (ductal non special type (NST) infiltrating carcinomas), but it could have difficulties with lobular carcinoma, in situ or patients who have undergone neo-adjuvant chemotherapy (NAC).

3.2. Virtual biopsies

Once the MCR modelling and PLS-DA classification abilities have been demonstrated, the next step is to select the best model and apply it for prediction. This model, as explained in section 2.3, is obtained by averaging the $\mathbf{B_{PLS}}$ coefficients of the 10 best models. It should be noted that the group of best models were compared in terms of the regression coefficients and the differences between them were found not relevant (statistically significant differences in ANOVA, p-values>0.05) (results not shown). Therefore, the average model is a good representation of the group of best models. Then, the corresponding regression parameters of the PLS-DA model associated to each biomarker are taken and applied to the external data set mentioned on section 2.1 (60). Some examples of histologically-confirmed tumor and healthy cases are shown in Fig. 8:

For prediction purposes, the best threshold for class separation is established in +2 as the prediction can take values up to 3. For this model, a value higher than 2 (in the breast region) can be considered high tumor probability and lower than 2 can be considered healthy tissue. As can be seen, the virtual biopsies are able to predict the tumor region with precision indicated in Fig. 8 with white arrows. Thus, the visual interpretation is clear and easy (with only one parametric map), supporting and helping the radiologists in their daily work. Posteriorly to this work, this model has been applied on a clinical multicentric evaluation to a more extended data base of 136 cases, yielding 96.2 % sensitivity and 84.7 % specificity with 23 and 4 FP/FN respectively on patient level (the results of the present work are on pixel level).

4. Conclusions

MCR-ALS imaging biomarkers were able to successfully extract the pure behaviors (associated to vascularization processes) from DCE-MRI sequences in breast cancer. Also, their ability to discriminate tumoral from healthy tissue has been demonstrated, showing average f-score of 0.9149 and average MCC of 0.8538 which are the best results obtained from MCR-ALS models for cancer discrimination till now. These results are better than the ones published for prostate cancer (f-score = 0.7857) [21] and others where only first generation pharmacokinetic models are considered (sensitivity, between 0.45 and 0.63; specificity, between 0.71 and 0.85) [22–24], AUC (0.7) [25]. Moreover, these results remark the relevance of MCR for extracting very useful information in situations where the DCE-MRI pharmacokinetic clinical biomarkers reach a computational limitation. Finally, new parametric maps known as "virtual biopsies" have been proposed as a clear and easy way of breast cancer detection, presented as a helping tool for radiologist's daily cancer diagnosis tasks.

CRediT authorship contribution statement

E. Aguado-Sarrió: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **J.M. Prats-Montalbán:** Conceptualization, Funding acquisition, Validation, Visualization, Writing – review & editing. **J. Camps-Herrero:** Investigation, Resources, Supervision, Validation, Writing – review & editing. **A. Ferrer:** Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

Table 4
Percentage of inclusion of the biomarkers from MCR-ALS in the PLS-DA model, regarding the variable selection method.

Biomarker	% of inclusion
C_1	96.73
C_2	99.73
C_3	98.53
RSS	97.93
B2	100
B1	99.33
B3	99.8

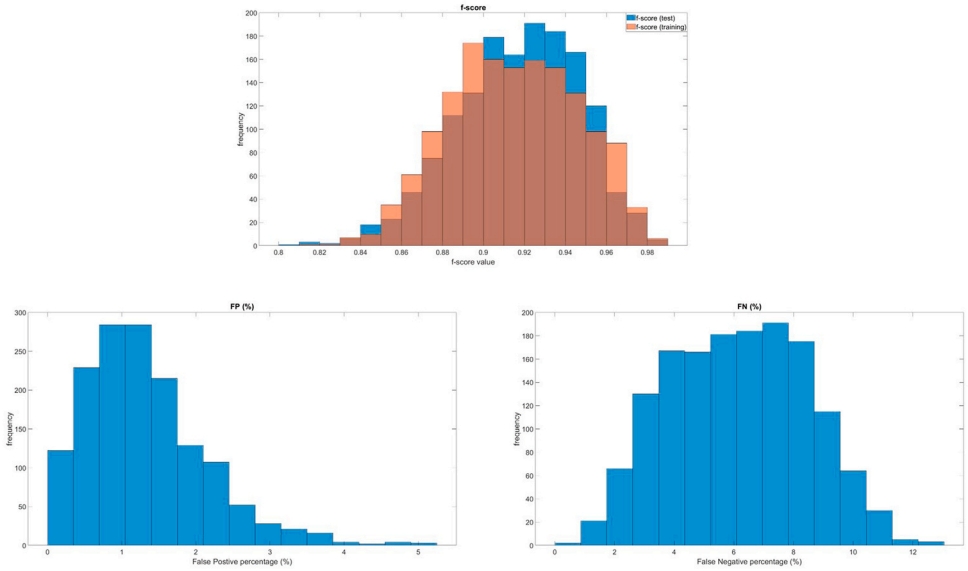


Fig. 7. Top: f-score (test) vs. f-scoreT (training), bottom left: %FP, bottom right: %FN. Calculated for the 1500 iterations of the PLS-DA iterative model.

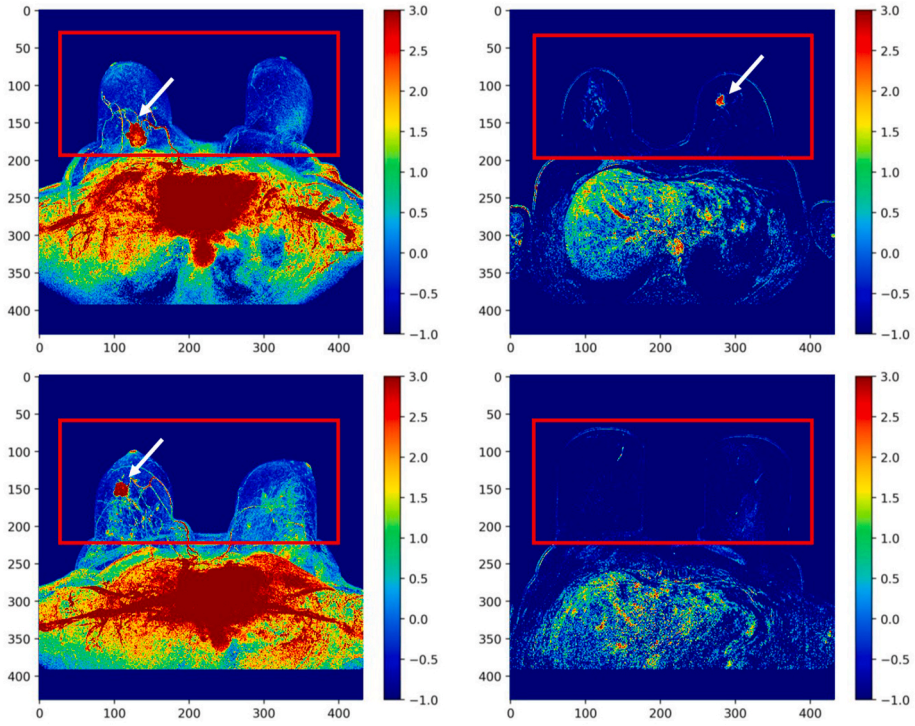


Fig. 8. Virtual biopsies examples calculated for external cases. Scale from −1 to 3. Values > 2 show high probability of breast cancer in the breast region. The region of the body in the lower part of the image (below the breast) is not considered for cancer prediction as it contains vascular regions like the heart or big arteries. The breast region is indicated by the red squared section. The white arrow indicates the location of the tumor lesion in the image. The last example (bottom right) shows a virtual biopsy map of a healthy case. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This research was supported by the Spanish Government (Science and Innovation Ministry) under the project PID2020-119262RB-I00, and by the Generalitat Valenciana under the project AICO/2021/111.

References

[1] S. Lei, R. Zheng, S. Zhang, S. Wang, R. Chen, K. Sun, H. Zeng, J. Zhou, W. Wei, Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020, Cancer Commun. 41 (11) (2021) 1183–1194.

[2] Cancer Today. International Agency for Resarch on Cancer. World Health Organization (https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf) .

[3] C. Allemani, T. Matsuda, V. Di Carlo, et al., CONCORD Working Group, Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries, Lancet 391 (10125) (2018) 10–23] -1075.

[4] A. Sheye, Diffusion MRI Methods for Improved Treatment Monitoring in Breast Cancer, University of California, Berkeley, 2009. T.J. Christmas, A.C. Thompson, M.J. Bailey, C.M. Corbishley, C. Fisher, M.O. Leach, D.P. Dearnaley, Dynamic contrast-enhanced MRI for prostate cancer localization, Br. J. Radiol. 82 (2009) 148–156.

[5] D.J. Collins, A.R. Padhani, Dynamic magnetic resonance imaging of tumor perfusion, IEEE Eng. Med. Biol. Mag. 23 (2004) 65–83.

[6] A.S.N. Jackson, S.A. Reinsberg, S.A. Sohaib, E.M. Charles-Edwards, S. Jhavar, T. J. Christmas, A.C. Thompson, M.J. Bailey, C.M. Corbishley, C. Fisher, M.O. Leach, D.P. Dearnaley, Dynamic contrast-enhanced MRI for prostate cancer localization, Br. J. Radiol. 82 (2009) 148–156.

[7] M.O. Leach, K.M. Brindle, J.L. Evelhoch, J.R. Grif ths, M.R. Horsman, A. Jackson, G.C. Jayson, I.R. Judson, M.V. Knopp, R.J. Maxwell, D. McIntyre, A.R. Padhani, P. Price, R. Rathbone, G.J. Rustin, P.S. Tofts, G.M. Tozer, W. Vennart, J. C. Waterton, S.R. Williams, P. Workman, The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations, Br. J. Cancer 92 (2005) 1599–1610.

[8] J.M. Prats-Montalbán, R. Sanz-Requena, L. Martí-Bonmatí, A. Ferrer, Prostate functional magnetic resonance image analysis using multivariate curve resolution methods, J. Chemom. 28 (2014) 672–680.

[9] J.M. Prats-Montalbán, E. Aguado-Sarrió, A. Ferrer, Multivariate curve resolution for magnetic resonance image analysis: applications in prostate cancer biomarkers development, in: "Resolving Spectral Mixtures, with Application from Ultrafast Spectroscopy to Super-resolution Imaging", Data Handling in Science and Technology, vol. 30, Elsevier, Amsterdam, 2016, pp. 519–550.

[10] M. Sjöström, S. Wold, B. Söderström, PLS discriminant plots, in: Proceedings of PARC in Practice, Amsterdam, June 19–21, 1985, Elsevier Science Publishers B.V., North- Holland, 1986.

[11] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.

[12] E.A. Morris, C.E. Comstock, C.H. Lee, et al., ACR BI-RADS® magnetic resonance imaging, in: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, American College of Radiology, Reston, VA, 2013.

[13] P. Geladi, H. Grahn, Multivariate Image Analysis, Wiley, Chichester, England, 1996.

[14] J.M. Prats-Montalbán, A. Ferrer, A. de Juan, Multivariate image analysis: a review with applications, Chemometr. Intell. Lab. Syst. 107 (2011) 1–23.

[15] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.

[16] M.J. Bruwer, J.F. MacGregor, D. Noseworthy, Dynamic contrast-enhanced MRI diagnosis in oncology via principal component analysis, J. Chemom. 22 (2008) 708–716.

[17] R. Tauler, A.K. Smilde, B.R. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58.

[18] R. Tauler, Multivariate curve resolution applied to second order data, Chemometr. Intell. Lab. Syst. 30 (1995) 133–146.

[19] M. Sjöström, S. Wold, B. Söderström, PLS discriminant plots, in: Proceedings of PARC in Practice, Amsterdam, June 19–21, 1985, Elsevier Science Publishers B.V., North-Holland, 1986.

[20] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochim. Biophys. Acta Protein Struct. 405 (2) (1975) 442–451.

[21] E. Aguado-Sarrió, J.M. Prats-Montalbán, Requena Sanz, G. García-Martí, L. Martí-Bonmatí, A. Ferrer, Biomarker comparison and selection for prostate cancer detection in dynamic contrast enhanced-magnetic resonance imaging (DCE-MRI), Chemometr. Intell. Lab. Syst. (2017).
[22] A.S. Jackson, S.A. Reinsberg, S.A. Sohaib, E.M. Charles-Edwards, S. Jhavar, T. J. Christmas, A.C. Thompson, M.J. Bailey, C.M. Corbishley, C. Fisher, M.O. Leach, D.P. Dearnaley, Dynamic contrast-enhanced MRI for prostate cancer localization, Br, J. Radiol. 82 (2009) 148–156.
[23] T. Franiel, L. Lüdemann, B. Rudolph, E. Lutterbeck, B. Hamm, D. Beyersdorff, Differentiation of prostate cancer from normal prostate tissue: role of hotspots in pharmacokinetic MRI and histologic evaluation, AJR Am. J. Roentgenol. 194 (2010) 675–681.
[24] R. Sanz-Requena, L. Martí-Bonmatí, R. Pérez-Martínez, G. García-Martí, Dynamic contrast-enhanced case-control analysis in 3T MRI of prostate cancer can help to characterize tumor aggressiveness, Eur. J. Radiol. 85 (2016) 2119–2126.
[25] E.K. Vos, G.J. Litjens, T. Kobus, T. Hambrock, C.A. Hulsbergen-van de Kaa, J. O. Barentsz, H.J. Huisman, T.W. Scheenen, Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced MRI at 3T, Eur. Urol. 64 (2013) 448–455.